Ethics and Information Technology (2006) 8:3–15
DOI 10.1007/s10676-006-9105-3

Designing databases that enhance people's privacy without hindering organizations

Towards informational self-determination

Thomas B. Hodel-Widmer
Center for Security Studies, Swiss Federal Institute of Technology (ETH Zurich) ETH Zentrum, LEH D 6, Zurich, 8092, Switzerland
E-mail: hodel@sipo.gess.ethz.ch;

Abstract. We argue that future database systems must provide autonomy for individuals for the privacy of data they manage. We propose a design for such a system, identify challenges and problems, and suggest some approaches to these. We enunciate the reasons for informational self-determination systems, which include legal, organizational and technical issues. Our main goal is to achieve a widely-accepted realistic and practical solution in order to ensure privacy for individuals in our future world, yet without hindering business and security.

Key words: database management systems, informational self-determination systems, Privacy Enhancing Technology (PET)

Privacy issues of privacy invasive technology

Identity-related technologies, location-based services and ambient intelligence technologies certainly risk the loss of anonymity and raise a number of privacy concerns.

Identity management systems could play a critical tool for the citizen, based on the growing number of services. A unique access tool creates an 'electronic' witness of a major part of the citizen's online life. Furthermore these technologies have to comply with the regulatory framework concerning privacy rights. To manage and control these data is very difficult, especially if the entities are not from the same regulatory territory.

Biometric data are sensitive and of a personal nature. Therefore even if it is forbidden by law, the risk of being disclosed to a third party is high. Numerous privacy concerns are raised within this topic. Biometric data fully identify a person and provide additional and sensitive information. Medical specificity can be found in fingerprints, iris image, and retina scan, for example. Iris scan and face recognition do not require contact therefore they are more risky for privacy, because they can be diffused or hidden in the local environment.

RFID tags can be accessed as well 'contact less' since they are also invisible. Therefore RFID tags raise specific privacy concerns such as user awareness and empowerment. RFID tags represent a sort of identity management system, as soon as the tags are linked to the owner of an object, thus defining the extent of privacy compromise.

In the near future, cellular system and WLAN technologies will diffuse mobile broadband services. Wireless communication increases privacy concerns regarding personal data, traffic data and location data. Negative consequences may arise for users when databases are mined.

Location based services raise important privacy questions. All citizens will have a shadow in the virtual world. Physical location and movements will be stored as personal, traffic and location data within the virtual world. Different parties are involved in the value chain of a location based service, therefore there is an even higher risk with regard to respecting and protecting privacy rules.

Monitoring and surveillance capabilities, using ambient intelligence, will emerge on a large scale.

Ethical reflection

As mentioned above, we are geared to the concept of autonomy, which has become a key notion in the Modern Era (since the 18th Century). In

consequence, autonomy functions as a guiding principle for one's action and as a normative idea to be followed in social and technological evolution The concept of autonomy thus serves as a standard for the ethical evaluation of individual action and of technological development. Of course, the intellectual content of it still needs clarification; yet it functions as an ethical authority with its critical, its directive and constructive, in short with its challenging potential. Thus the concept of autonomy is a true representative of modern consciousness, which insists on the humans' responsibility of their moral and legal rights, as well as on the respect for possible self-determination.

Autonomy as self-determination

In today's interdisciplinary discussion of autonomy, the second element of what we just mentioned takes the lead, i.e. self-determination in a moral, political, and legal sense. Here the autonomy principle calls for the respect of a person's private sphere as well as of its external sphere of action, as far as it depends on the person's will. So far, the autonomy principle is a defensive principle; it is meant to regulate external actions, limitations, also in the form of networks, infringements, yet also co-operation, as experienced by the individual human being. Thus the concerned person's right of self-determination must not be limited without good reason, nor against his or her will. However, it is not possible to derive from the autonomy principle – fundamental as it may be – any clear instruction for or any precise evaluation of a specific action. For this requires the immediate confrontation with the factual context. And yet, accepting autonomy as a human potential and a justified claim entails clear consequences. I. Kant has based his concept of autonomy on the conviction that normative obligation presupposes the rational subject's pledge to universalized norms he or she has formulated him- or herself. He thus laid the ground for our claim to self-determination and the ensuing personal responsibility. We may realize and conserve them both, but only as far as we succeed in controlling, refusing, accepting, or modifying external influences (heteronomy) on parts or the whole of our personal identity.

Informational self-determination

With regard to the so-called "informational self-determination", we may say that the principle of autonomy entails the right of monitoring the conditions under which personalized information is acquired and used (ALLEA 2002). This is admittedly a general, yet still a substantial principle, which needs a circumspect political and legal – and not in the least, technical – concretization. It goes without saying that the current trend will lead to more personal data being generated and used to an ever greater extent. As a result, there will be a greater risk to "informational self-determination".

This problem must be addressed, because all parts of our society are morally and legally concerned when it comes to guaranteeing and protecting "informational self-determination". This is a target value that must not be lost sight of and one that must be articulated repeatedly, particularly when this target value becomes just one element in a process of weighing goods and of due consideration. The autonomous informational self-determination is not an absolute value; it may lose some of is weight in the light of higher and well founded interests (e.g. in criminal prosecutions). In view of certain interests that sometimes compete with each other, such as among individual rights and protection, and public concerns, this principle must be appropriately implemented; especially when considering the consequences of disclosing personal data. This applies to the rules derived directly from the principle of informational self-determination (Privacy and the National Information Structure 1995).

Basic rules

(a) As a matter of principle, an individual should have the right to decide upon the conditions of the generation and use of his/her own relevant personal data; directly following from which, that any personal information resulting from this particular data is collected, disclosed, and made use of so as to be fully respect the individual's data protection. This concerns "Informational Privacy". The extent of this "Informational Privacy" protection should be defined and guaranteed through the best possible means. In most cases, this may be achieved when the concerned parties and the users of the particular information reach an agreement on exactly how the personal information should be disclosed and used. To reach this goal, it is indispensable that the society guarantees a determined fundamental level of data protection rules – with appropriate sanctions, in the event of misuse.

(b) When personal Information is stored, the concerned individual may rightfully claim that the respective information is correct, current, and complete in relation to the purpose for which it is intended and used. Therefore, the data must exhibit that it is of reliable quality (Principle of

Quality). Furthermore, the users of personal data must in principle furnish information on why such information have been collected, what they are used for and what measures have been taken in order to guarantee the confidentiality and the quality of the particular information (Enlightenment Principle or Notice Principle). This principle serves to guarantee that the individual's information is sufficient (and presented in a comprehensible manner) so as to allow him or her to give his or her informed consent (Principle of Informed Consent).

(c) Users of the particular information must undertake appropriate technical and organizational measures in order to guarantee the confidentiality and quality (Protection Principle) of personal related data. If such data are processed in a network-type environment, they are under high risk with regard to a third party's unauthorized access, unauthorized transmission, unauthorized modification of data, etc. Appropriate protection requires a discriminating, diversified, and a continuously verified strategy.

(d) The ''Principle of Fairness'' states that those who are accruing advantages based on their rights are not to be doing so as to the disadvantage of others (J. Rawls). In the context of ''Informational Self-Determination'', this means that personal information is not to be used in opposition to the concerned individual's expectations on how these data should be – with the exception of well justified public interests permitting such use. In this respect, ''fairness'' means that the users of the particular information have to accept the expectation as developed from the concerned individual's point of view, in order to determine to which extent the data may be used. The goal consists in guaranteeing that those who are concerned and the data users achieve a balanced understanding on what shall be the adequate precautionary measures for suitable, specific use of personal related data that should be taken.

Security and privacy

Balancing security and privacy in the information society will be a tough task. Respecting somebody's private life has to be weighed up against issues of national security, public safety, economic wellbeing, prevention of disorder and crime, protection of health and rights and freedom of others (Maghiros et al. 2003, Walters 2001). It is impossible to make a prediction as to which side the future will lie on, but the risk of losing privacy, the ''right to be let alone'', ''the right to select what personal information about me is known to what people'', in the information society is rather high (Westin 1967; Warren and Brandeis 1890).

From our point of view, citizens will lose their entire privacy if nothing is done against current developments. To strengthen privacy and security, actions on legal, organizational and technical issues are required (Lessig 1999). These three elements are included in our approach to privacy-enhanced database systems which we explain in Sections ''The concept of privacy revisited'' and ''Founding principles for informational self-determination database systems''.

State-of-the-art privacy enhancing technology

The term Privacy Enhancing Technology (PET) originated in the mid-nineties from a study that investigated technological measures to curb the use of identifying data in information systems (Registratiekamer 1995). Nowadays the term PET is widely used, and refers to technologies which aim to eliminate the use of personal data in information systems or to restore the user's control over the revelation of personal data (Burkert 1997). In a wider sense, one could say that the term PET represents all technologies which pertain to protecting an individual's privacy.

Many privacy enhancing technologies aim to allow anonymous transactions and anonymous communication in the Internet. While this is clearly the most effective approach to avoid the creation of personal data, it remains to be seen whether service providers are willing to embrace these technologies. The approach of enterprise-level privacy policies promises to guarantee that enterprises do indeed process data according to their declared policy. However none of the technologies we know offer a combination of legal and technical means to restore users control over personal data. New approaches need to be found that let users regain control of personal data stored in a multitude of databases.

The concept of privacy revisited

More than 100 years ago, Warren and Brandeis wrote the landmark paper ''The Right to Privacy'', published in the Harvard Law Review in 1890 (Warren and Brandeis 1890). They defined privacy as ''the right to be let alone'' and argued that legislation should give this right to every individual: ''Political, social, and economic changes entail the recognition of new rights.'' (Warren and Brandeis 1890) In the twentieth century, many legal scholars and

philosophers have attempted to define the concept of privacy (Gormley and Brandeis 1992). However, it is impossible to come up with a universally valid definition of privacy as the concept depends on social aspects, cultural values and the legal framework. The issues of privacy are "fundamentally matters of values, interests and power" (Gellmann 1998, p. 194).

An implication of privacy as an interest is that it has to be balanced against other competing interests. People's interest in their own privacy may conflict with the interests of other people or organizations (Etzioni 1999). The concept of privacy does not apply to information only. Privacy rights have a long tradition and are implemented in many fields (Rosenberg 1992):

- Territorial privacy: Protects the physical surroundings of a person, i.e. in a domestic or other environment.
- Bodily privacy: Protects the physical integrity of a person against undue interference (e.g. physical searches, DNA testing).
- Communication privacy: Protects the personal communication of a person against monitoring by other people or organizations.
- Informational privacy: The right of a person to control what data about him or her can be gathered, processed and disseminated.

In the context of information systems, privacy has naturally been defined as informational privacy. This restriction makes sense as an information system usually does not affect territorial or bodily privacy (with the exception of robotics applications or some ubiquitous computing devices, which are outside the scope of this paper).

A very common and well-accepted definition of informational privacy is the one given by Alan Westin in his classical work on privacy. Westin defines informational privacy as "[t]he claim of individuals, groups and institutions to determine for themselves when, how and to what extent information about them is communicated to others" (Westin 1967).

At the heart of the notion of informational privacy lies the understanding that certain information about a person is not public but rather private, however it is not possible to give a precise definition as to which data fall into which category. Such a notion depends on cultural understanding and personal views. Informational privacy is, just like other forms of privacy, the interest of an individual that may compete with the interests of other parties.

With the widespread use of information systems, the understanding of privacy shifts towards an interpretation of privacy as the right to informational self-determination. An individual should have the right to control the release and dissemination of personal data as well as the context the data are going to be used in, to the greatest possible extent. In addition to Alan Westin's definition of informational privacy, we state that in general informational privacy and the measures to protect it should address:

- the release and dissemination of personal data (purpose),
- the right to remain unidentified (anonymous) when we choose to,
- the protection of highly sensitive data in electronic systems,
- the latent danger of tracking and logging of users and their activities,
- the right to be let alone,
- the right to live without the threat of constant surveillance by electronic means.

We claim that the advent of new technologies poses a threat to the citizen's privacy. The fact that computers are becoming ubiquitous – and that information technology is becoming more and more a part of our daily lives – leads to an erosion of informational privacy. An awareness of privacy problems must therefore be created urgently. We maintain that any technology able to enhance privacy is thus worth discussing. We see our paper as a contribution to the discussion on privacy issues and aim to point out new directions in which technology and legal frameworks may be developed in order to work towards offsetting the negative effects that information technology has on privacy. The next section further motivates the principles of participation and transparency. Transparency and participation are considered in the context of the data protection tradition. Both are discussed in the context of private as well as public sector data processing. We consider how these two principles are implemented by our architecture and explain why the architecture leads to more transparency and better participation as compared to most of today's data processing systems.

Transparency and participation in privacy legislation

When the interest in informational privacy began to increase due to the widespread use of information technology, legislative bodies began addressing the problem in the 1970s. The first modern data protection act was adopted by the German State of Hessen, the first national law by Sweden in 1974. A very influential piece of data protection legislation is the US Privacy Act. The act was passed by the Congress in 1974, thereby acknowledging that the rapid development of information systems posed a threat to

personal privacy. Although the Data Protection Act was not very successful in the US, it found much attention abroad. This resulted in the fact that many elements of this policy can be found in data protection laws of other countries.

The US Privacy Act was crafted after the work of an advisory committee, which had established the notion of "fair informational practices". These practices are based on work by Alan Westin and turned out to be very influential in shaping data protection legislation around the world. Westin stated eight important principles for fair information processing (Westin 1967). These principles have been also incorporated into the Organization for Economic Cooperation and Development guideline on data protection of 1980 (OECD 1980) and the EU directive 95/46/EC on the protection of individuals with regard to the processing of personal data (The European Parliament 1995).

One of those principles is the principle of openness and transparency. It states that there should be a general policy of openness guiding collections of personal data. Especially, there should not exist any secret data collections. Means of establishing the existence and nature of collections, the main purposes of their use as well as the identity of the data controller should generally be known. Another important principle is the principle of individual participation: Individuals should have the right to request information from a controller as to whether a collection contains data about them. Requests should be answered within a reasonable period of time and at a reasonable price. Furthermore, individuals should have the right to have records rectified, completed or erased where appropriate (i.e. in the case of incorrect or illegally stored data).

Transparency and participation in e-commerce data processing

Various surveys have shown that privacy is a substantial concern on the Internet, particularly in e-commerce transactions (Ackerman et al. 1999). Users are obliged to divulge personal data in almost every transaction, and in so doing, leave traces each time such a transaction is carried out. In most business relationships, users have neither insight into what data the other party collects nor do they have access to these data.

For e-commerce purposes, P3P is slowly gaining popularity. This standard, however, only addresses privacy declarations. The use of P3P does not lead to any form of participation or to a much enhanced transparency. There are very few companies who allow users to see their personal data and to control how this data is to be used. An approach such as EPAL is therefore a step in the right direction. The use of EPAL guarantees that data are processed in accordance with specified polices. Our approach specifically aims at heightening transparency and participation in data processing. We thus conclude that in the domain of e-commerce, participation and transparency in data collections are the exception rather than the rule. An approach such as the one presented in this paper can help to make data collections more transparent and to give users more participative power. We shall propose that portals should be operated which give individuals access to the audit data (see Section "Participation") that is stored about them and thus increase both transparency and participation.

Transparency in e-government data processing

Informational privacy is an especially important issue in e-government. The data that are processed in e-government environments are often of a much more sensitive nature than the data processed in the domain of electronic business (Joshi et al. 2002). People are increasingly concerned about privacy issues related to e-government, and tend to feel the same way about citizen cards (BBC News 2003). Although information and communication technology provide notable opportunities for reshaping the relationship between government and stakeholders and for creating more efficiency in bureaucratic systems, it also creates significant security and privacy challenges. Data in governmental databases contain highly sensitive data such as social security numbers, information related to individual taxation, data concerning religious beliefs, criminal records, demographic information and medical records. Furthermore, governmental bodies process high volumes of data. They are empowered by public law to collect data on citizens and can enforce their right to do so. Governments thus have the potential to accumulate large data collections, which may create potential conflicts with the citizen's interest in informational privacy (Schweizer and Burkert 1996). Given these facts, it is even more desirable that citizens know what data governmental administration keeps about them.

Administrative cultures and procedures in Europe vary, and so do the views on the sensitivity of data. Religious affiliation is considered a very sensitive issue in the Netherlands and in Greece, while inhabitants of Finland are very sensitive about data that relates to the gender of a person. Many other examples can be found illustrating the differences that exist with regard to the sensitivity of data.

We feel that there is still a general lack of transparency and participation in governmental data collections. In most European countries, citizens do not have the right to access their own data in governmental data collections. An exception is Sweden: All data that is collected by the state is deemed public. As a consequence, any citizen has the right to see e.g. his neighbor's tax declaration. Another fairly advanced country (with regard to participation) is the Netherlands: Here it is currently being discussed if citizens should have access to their own data in all governmental data collections. Yet in most European countries citizens do not automatically get access to their own public records.

Conclusion

With the widespread use of information systems, the focus on privacy shifts towards an understanding of privacy as the right to informational self-determination. An individual should have the right to control the collection, the release and dissemination of personal data as well as the context in which the data is going to be used, to the greatest possible extent.

Founding principles for informational self-determination database systems

Privacy enhancement can be understood as an increase in the control each customer has regarding personal data which is shared with organizations. In this section, we introduce our concept for privacy enhancement and point out the key principles on which our system design is based.

Our founding principles are motivated by our ethical reflection described in Section ''Ethical reflection''. These principles are rooted in existing data protection laws. They articulate what it means for a personal data collection system to responsibly manage private information. We argue for the following six newly interpreted principles that should complement the several privacy regulations which already exist. Indeed some of the principles are related but not similar to Westin and Agrawal (Westin 1967; Agrawal et al. 2002).

Consent: People know when their personal data are stored and have to give their consent prior to storage.

Purpose: People affected must have the possibility to specify the purpose and usage of their data.

Separation: Personal data have to be stored separately from business data.

Audit: Transactions involving personal data must be recorded in transactional logs. Persons affected can then follow executed transactions and retrace usage of their personal data.

Participation: People affected have access to their personal data, its usage and purpose specification. They can choose if, when and how to manage their personal data.

Ease of use: People affected have the choice to bundle access to personal and audit data through portals and can define automatically applied patterns.

In comparison with Westin and Agrawal, principles such as ''limited collection'', ''limited use'' and ''limited retention'' are not requested within our approach, however each individual is free applying his own set of principles. Within our approach, the ''consent'' principle is enforced by law and is strictly connected to the ''purpose specification'' principle, which is supported by technology. This infrastructure is expanded in such a way that each individual knows all his or her data sources. This makes principles like ''limited retention'', ''openness'' and ''compliance'' traceable, so that mistreatments of the data-protection law can be investigated. Principles such as ''accuracy'' or ''safety'' are essential requirements and, as such, will not be mentioned again.

Consent

Nowadays almost any transaction is recorded. As long as no exact identification of a specific person can be made by using these data, no privacy issues are involved and there is no need for us to care about it. As soon as these data are linked to personal data, however, privacy could be jeopardized as described in Section''Privacy issues of privacy invasive technology''.

The first principle is that people, whose private data are stored, must give their consent for this storage, and the specified organization is obliged to inform these individuals ''where and what'' data are stored. In most cases, people do not remember which companies store their data; they often don not have any possibility to know this because in many cases they are completely unaware of such a data collection.

Personal data can be used for evaluations and for marketing purposes. It may be sold to other companies without the customer's consent or knowledge, and such data could even be stolen. Generally people do not pay attention to who manages or what happens with their data, but as soon as they are harassed with spam, telemarketing calls or advertising mails they want to know how this problem has arisen. On the other hand, it is important that organizations are not able to refuse services to any individual on the grounds of a possible risk. Excluding customers from

setting up a life insurance policy, denying access to buildings or generally concealing information are just a few examples of this. The importance of giving customers more information about data storage and the necessity of the customer's consent for further usage of that data is evident. At the same time, organizations gain competitiveness while data management transparency is offered to customers.

Purpose

The first principle illustrates the importance of customers being informed as to where and what personal data is stored. We outline why it is important to specify the purpose as to how personal data can be used.

Personal data can be used for different purposes and it is often used against people's intentions. This data-misuse problem can be solved if organizations put the people affected in a position from which they can influence further data management. Each organization defines its own purposes which determine the intended use of personal data. Individuals are then able to decide how these settings should be applied to their personal data. For example, a purpose specification may be to receive special offers by e-mail. Organizations can distinguish themselves from competitors and at the same time enhance trust and confidence in their services. This method of participation naturally varies from organization to organization. The only exceptions when people's personal data is passed on without their consent are defined by legal regulations or occur during criminal investigations.

Separation

An area urgently requiring more attention with respect to privacy and security is the stage at which business data is separated from personal data. During such a separation, business data, which contains sensitive information (e.g. about executed transactions), can be used for data mining without any need for the person's consent. Only an identifier indicates that these data belong to a specific person, so the data are anonymous as long as no connection to personal data can be made. As soon as personal data are requested for a specific purpose by linking to these data, this process must be permitted by the person affected and subsequently, recorded in the audit trail.

Audit

Both people and organizations must have the possibility to understand and detect unauthorized uses of personal data. This leads us to the need for audit information where all executed transactions which accessed personal data can be traced. Such information should contain all of the following: Who had, when and with which purpose, access to what kind of personal data? This knowledge provides more security to individuals and organizations. This audit information simultaneously supports data protection and helps to minimize fraud. Usually, these data are stored on the organizational side, but should be readily accessible to the persons affected.

Participation

While discussing the principles above, we saw why it is so important for people to manage and control the usage of their data. On the one hand, customers must be informed about further utilization of personal data, and on the other hand, they must be able to give their consent for any usage purpose.

To fulfill these requirements, customers need access to personal data being stored on the organizational side. This participation can be realized in different ways, e.g. by telephone, forms or internet.

Ease of use

A possibility of accessing personal data is realized via web portals. The central idea is to aggregate the information shared with all the organizations we are dealing with, and to create one personal portal. This provides people with a better overview and ensures that organizations know where users are managing their data and that they are informed of any changes. The resulting benefit for organizations is improved customer contact, enhanced trustworthiness and a higher level of confidence.

This kind of information aggregation results in a possible security gap. Each person can minimize this problem by depositing their personal data on different web portals. Each portal is physically separated, certificated and protected by a password.

This solution encompasses good standards, open interfaces and the possibility for organizations to buy these systems out of the box, its main objective being to enhance the ease of use by offering standardized interfaces and by always adhering to the security requirements.

Design

In this section, we discuss the design aspect. We study a scenario and illustrate the idea of purpose specification with the help of two examples. Furthermore,

we outline the structure to indicate the direction in which the set-up of such databases could be preceded. However, it is not a full implementation guide.

A use scenario

Avantara and Belios are two on-line booksellers who want to enhance customers' confidence in their company by implementing an autonomic database system. The main idea is to provide a service giving customers the possibility to define what happens after personal data is entrusted to their companies. Basically, customers set purposes for their personal data usage. During the process in which business data is separated from a customer's personal data, these anonymous business data can be used for data mining and data analysis. References from business to personal data always need a customers' consent.

Additionally, customers are able to see and verify all executed transactions in a transactional list (audit trail), which is automatically updated each time the personal data is accessed.

In this section, we look at examples revealing how the two booksellers handle this requirement and what purpose specifications they define.

Purpose specification Belios

Avantara and Belios must observe legal regulations and inform customers about these exceptions. For example, in the case of criminal investigations, personal data may be handed over to public agencies without the customer's consent. Avantara and Belios have different opinions about how much information and customer's cooperation is necessary. Belios defines only a few settings (see Figure 1) for purpose specifications of personal data, and only asks general questions, for example, if the customer would like to receive advertisements.

Purpose specification Avantara

Avantara, on the other hand, gives customers various possibilities to define purpose specifications regarding the use of their personal information. For instance, Avantara assumes that customers have preferences as to which information should come via which channel. Hence, Avantara offers various channels for communication and makes distinctions between private and business phone numbers. Furthermore, customers can classify how they prefer to be contacted. These options are contracted under the tab "Contact". Under "Order", general order properties are defined, such as whether or not customers wish to be informed about their order status. Other companies and individuals are also employed to perform functions on Avantara's behalf. Examples include fulfilling orders

Figure 1. Privacy control settings for customers of Belios.

and delivering packages, sending postal mail and e-mails, etc. They require access to personal information which is necessary in order to perform their functions, but they are not permitted to use it for any other purpose. Avantara guarantees that business or personal data is never passed on to third parties without the customer's prior consent, and that customers are always asked if data may be used for purposes other than those defined at the beginning. For customers who do not want to answer each single question under the "Defaults" tab, Avantara defines settings-categories for data usage. The data usage allowance can be set on "Minimum" or "Maximum". Last but not least, Avantara gives customers the chance to define the intensity of advertisement. These predefinitions are visualized in Figure 2.

Alice and Bob are looking for a skilled online bookseller, whereby Avantara and Belios are shortlisted. Alice is a privacy fundamentalist who normally doesn't want companies to retain any information once her purchase transaction is complete. However, she is willing to commit her personal data in order to receive some specific information if she can be certain that her data will be handled confidentially and only for the chosen purposes. For this reason, Alice decides to buy her books at Avantara since there she has the best overview of her personal data usage. Bob, in contrast, is a privacy pragmatist. He appreciates the convenience of only having to provide his e-mail and postal address once when registering with organizations. He likes to receive new recommendations, but does not want to be part of purchase circles. He also chooses Avantara but his reasons are different from

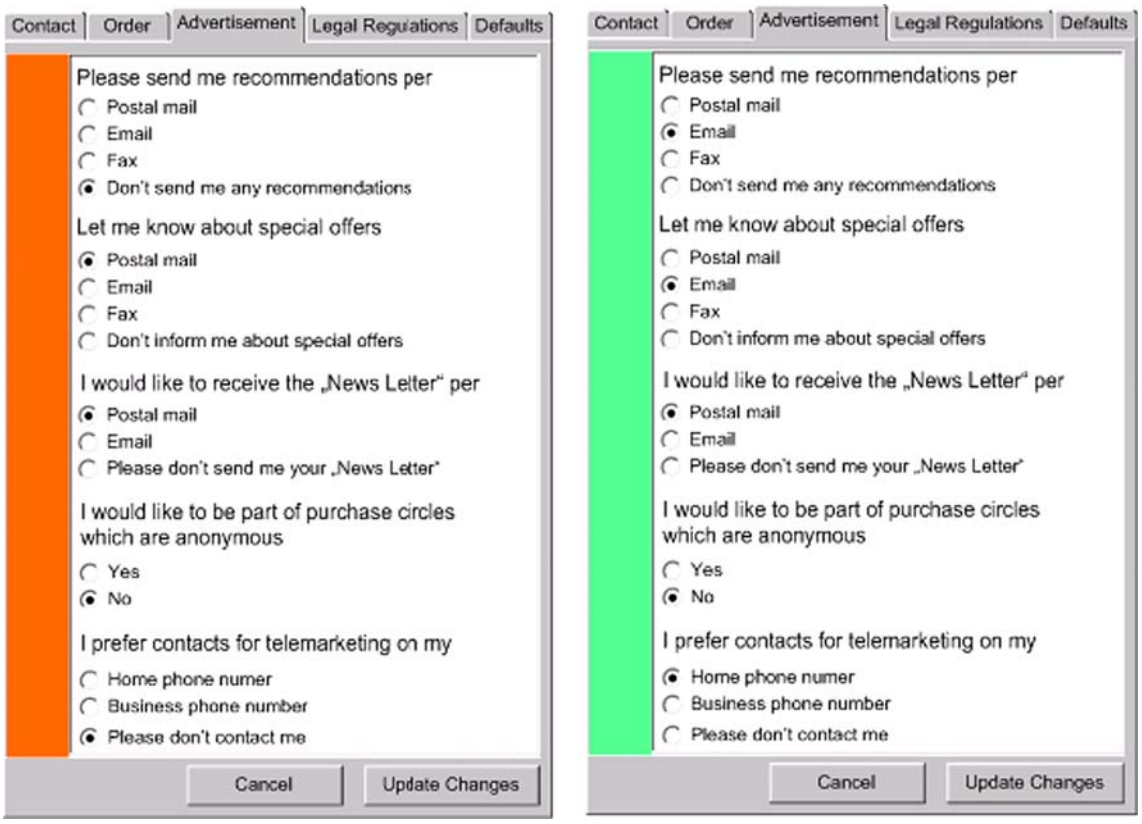

Figure 2. Privacy control settings of Alice and privacy control settings of Bob.

Alice's. The different "Privacy Control Settings" of Alice and Bob are illustrated in Figure 2. (compare also Agrawal et al. 2002). Trent is Avantara's privacy officer. He is responsible that the information system complies with the company's privacy policies. Mallory is an employee and he has questionable ethics.

Architecture

Finally, we present the architecture of an autonomic database. Central to the design is the active participation of customers in providing specific information within the organizational systems (compare Figure 3).

Components

Customer Data Requestor is responsible for opening a communication channel to the Request Handling Agent, which is located on the Customers Data System side.

Request Handling Agent only accepts properly formulated requests from the corresponding Customer Data Requestor.

Privacy Settings Rule Model covers rules which determine for which purposes customers' personal data can be accessed. These rules are constituted in the Privacy Control Settings. Trent designs these privacy definitions with regards to the company's privacy policy. For instance, he determines the purposes as to when a customer's e-mail address can be used.

Rule Compliance Validator examines whether or not a personal data request complies with the Privacy Control Settings of each user.

Access Control takes care of accesses before and during query execution. Access Control is carried out on both the Business and Personal Data Identification System.

Query Intrusion Detection checks the accuracy of accesses after the queries by comparing the access with the usual access patterns for queries with that purpose and by that user. For example, Mallory decides to steal all e-mail addresses of Avantara's registered users and to sell them to Avantara's competitors. Normally, customers' e-mail addresses can only be accessed for sending them recommendations or offers, or to enable order status tracking etc., as defined in the Privacy Settings Rule Model. Before the query results are returned, the Query Intrusion Detection matches these queries with the usual access patterns and detects the fraud.

Audit Trail records all possible queries for privacy audits and addresses challenges regarding compliance. Furthermore, this is where the customer's

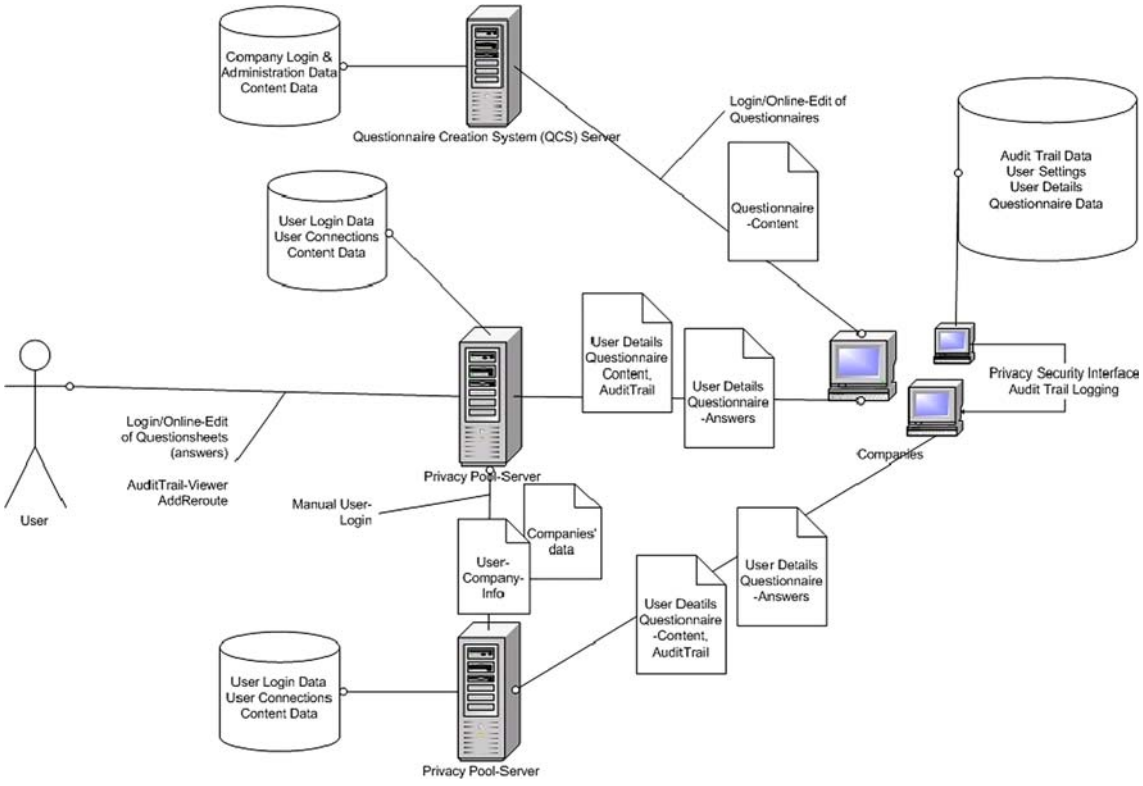

Figure 3. Architecture: collaboration model.

personal preferences as well as any changes to the Privacy Control Settings are maintained. Since customers have access to audit information, they are in a position to view all transactions and to detect any fraud.

Privacy policy

Figure 4 illustrates the separation of customers' personal and business data. The privacy policies of the two systems therefore differ in certain aspects, as explained in the following section.

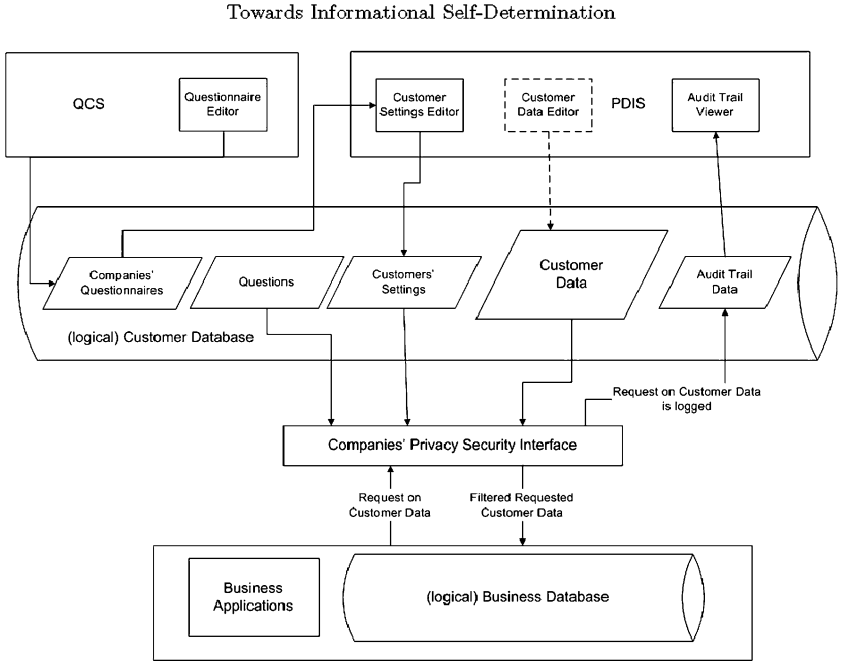

Figure 4. Security interface – positioning.

Authorized users and applications of the Business Data System are specified in the privacy policy. These are the set of Avantara's employees and applications who, or respectively which, can access particular information. The anonymous business data is accessible for purposes such as data maintenance, data mining and data analysis. As a result of the data separation, Avantara does not require a customer's personal information for most data mining and analysis activities – that is, not until Avantara addresses its customers directly.

The privacy policy for the Personal Data Identification System is more sophisticated and consists of three main parts.

Authorized users are a group of employees, customers and applications. Employees and applications access this data for maintenance purposes only. Customers, in comparison, access the Privacy Control Settings to assign their preferences and restrictions with regard to data usage. Moreover, customers access Audit Trail information to view and verify the suitability of the use of their personal data. Returning to our example case, Mallory is employed by Avantara to maintain customers' business data, therefore he has no authorization to access customers' personal data.

Rule Mechanism Privacy rules are defined in the Privacy Settings Rule Model. This model covers rules which determine the general purposes for which customers' personal data can be accessed. The Rule Compliance Validator checks customers' Privacy Control Settings to examine if specific accesses should be allowed.

Request/Reply Mechanism: The only way of connecting anonymous business data to customers' personal data is via a communication channel between the Customer Data Requestor and the Request Handling Agent. The Customer Data Requestor asks for information from the Request Handling Agent, which handles these requests and sends back a reply verified by the rule mechanism.

Queries

Avantara decides to launch a new marketing promotion, and therefore selects 500 records from the Business Data, with the intention of sending these customers specific recommendations by post or by e-mail. In order to do so, Avantara needs to access the Personal Data Identification System where customers' addresses are stored. The access from the Business Data System to the Customers Data System is only possible via a controlled channel. All queries for customers' personal data are first sent to the Customer Data Requestor. The Customer Data Requestor forwards these queries to the Request Handling Agent, which is located in the Personal Data Identification System (see Figure 4).

The Request Handling Agent passes on all properly formulated queries it receives to the Rule Compliance Validator. The query for customers' postal or e-mail addresses with the purpose "recommendation" was sent by an authorized employee at Avantara. The Rule Compliance Validator now checks, in accordance with the Privacy Settings Rule Model, if this query can be accepted. After this commitment, customers' Privacy Control Settings are checked. Alice stipulated in her Privacy Control Settings that she doesn't want to receive any recommendation whilst Bob would like to be sent recommendations by e-mail. Therefore only Bob's e-mail address is sent back to the Customer Data Requestor.

Let's suppose that Alice unexpectedly receives a recommendation from Avantara, despite having told them that she doesn't want this. Since Alice has access to the Audit Info where all transactions are recorded, she can verify the permission of the received e-mail and complain to Avantara about the mistreatment of her personal data (compare Figure 5).

New challenges

Now we describe some interesting problems which we identified in our principles and design. This list is by no means complete; its purpose is to initiate discussions. We use intentionally the same structure as in Section "Founding principles for informational self-determination database systems" to visualize the connection.

Consent

The cornerstone for informational self-determination database systems would be a new international data protection law requesting the explicit consent of a person before personal data can be stored. Furthermore, the law stipulates that this person must have access to their data, to specify purposes and to control audit information.

Within this law, several questions are raised. There will be a certain amount of administrative work and it will not always be clear how to set up the process. For instance, the user must first give his/her consent, before his/her personal data is stored, and not the other way around. How can organizations which do not care about this law be identified? Are normal individuals qualified to handle their personal data? Should one instruct a company specialized for this purpose?

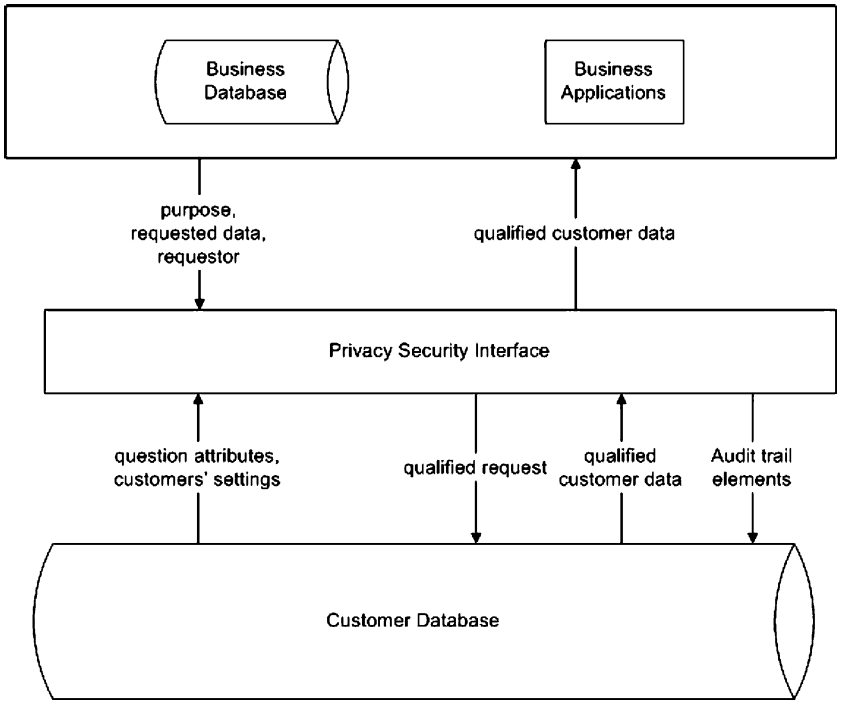

Figure 5. Flowchart: privacy security interface.

However – and this is a crucial point – at least one person knows which databases store information about him/her.

Purpose

At a first glance, purpose specification may appear easy. However selecting what kind of usage from personal data a person allows depends heavily on the way in which this can be achieved and on how these usages can be presented and categorized. No one is willing to spend several minutes specifying purposes; therefore, a low amount of fixed categorizations has to be defined in which each category includes several purpose specifications. Then, people can choose to make settings either only on the category level and/or for each purpose. The categorization must also be independent of the branch or industry. To set-up, define and become widely accepted, such a general categorization of purposes is essential and its development may be a tough task.

Separation

Business data and personal data are often already separated in large-sized companies. Different applications use these data. On the other hand, in small and middle-sized companies these data are normally stored together and are only used by one main application. A physical or logical separation is necessary according to the principle of separation. This makes any IT-architecture more complicated. In addition, the architecture has to be extended with a strong identification functionality. To increase trust and confidentiality, the "Personal Data Identification System" (see Figure 4) should be certified by a third party.

Audit

Generating audit trails that are in the hands of the people affected could provide a strong and powerful tool for protecting privacy. First of all, these audit trails can be investigated by the organizations themselves in order to detect internal misuse. Secondly, each person can scan these data and convince himself/herself in compliance with the audit trail of his/her personal data, or in the case of misuse, can place a complaint. Last but not least, a person can engage external software agents to monitor his/her audit information so that he/she can be automatically informed if a violation is detected. Within this scenario, three main questions arise. How can an individual set up his/her complaint and who will receive this message? What kind of competence or interest could such a "compliance office" persecute? What kind of consequences may occur for the principal offender? Furthermore, "Rule Compliance Validator" agents activated by the customer represent several security and privacy risks, despite being convenient for the customer.

Participation

Participation requests a certain kind of connection to the control equipment of the purpose specification and audit information. This communication and requested identification must be secure. Misuse cannot be tolerated.

Ease of use

We propose a hybrid solution. Each person can decide how centralized he/she would like to treat his/her personal data. A centralized system is quicker and easier to handle but encompasses more privacy risks than a decentralized system; however they could both provide a higher level of security. A centralized system is a far more attractive target for illegal transactions, because full data profiles related to specific users are available. The system's structure should at least be digitally secured against possible misuse and should guarantee the respect of a citizen's privacy.

Closing remarks

Organizations collect large amounts of personal data about their customers. Even though they promise privacy to their customers by means of privacy statements, there is no methodology to enforce these promises throughout and across multiple organizations. This paper illustrates the way in which the informational self-determination database systems can be achieved; it also defines legal and organizational principles as well as technical privacy-enabled database management systems for increasing personal privacy. Inspired by Kant, we present a vision of a system that minds the privacy of the data which it manages. Its comprehensive privacy-specific approach expresses how individuals regain control over their own personal data. From a user's point of view, this will tend towards the end of heteronomic database systems. Our approach offers a realistic, practical and pragmatic solution for enhancing individual privacy, without hindering business, organizations and national security doing their job.

References

M.S. Ackerman, L.F. Cranor, et al. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In S. Feldman and M. Wellman, editors, *Proceedings of the 1st ACM Conference on Electronic Commerce*, pp. 1–8. ACM Press, Denver CO, 1999.

R. Agrawal, J. Kiernan, et al. Hippocratic Databases. In *Proceedings of VLDB 2002*, pp. 143–154. Morgan Kaufmann, Hong Kong, China, 2002.

ALLEA, the European Federation of National Academies of Arts and Sciences. *Privacy Protection in the Information Society, All European Academies*, Amsterdam (ALLEA) 2002.

BBC News. *Public Oppose ID Card Scheme, 2003*. Available from: http://news.bbc.co.uk/2/hi/technology/3004376.stm, last visit: September 2004.

H. Burkert. Privacy-Enhancing Technologies: Typology, Critique, Vision. In P.E. Agre and M. Rotenberg, editors, Technology and Privacy – The New Landscape, pp. 125–142. MIT Press, Cambridge MA, 1997.

A. Etzioni, *The Limits of Privacy*. Basic Books, New York, 1999.

R. Gellman. Does Privacy Work? In P.E. Agre and M. Rotenberg, editors, Technology and Privacy, pp. 193–218. MIT Press, Cambridge MA, 1998.

K. Gormley. One Hundred Years of Privacy. *Wisconsin Law Review*, 1335–1441, 1992.

J.B.D. Joshi and A. Ghafoor et al., Security and Privacy Challenges of Digital Government. In W.J. McIver and A.K. Elmagarmid, editors, *Advances in Digital Government: Technology, Human Factors, and Policy*, pp. 121–136. Kluwer, Boston and Dordrecht, 2002.

L. Lessig, *Code and Other Laws of Cyberspace*. Basic Books, New York, 1999.

I. Maghiros and C. Centeno et al., *Security and Privacy for the Citizen in the Post-September 11 Digital Age: A Prospective*. IPTS Publications, Overview, Sevilla, 2003.

OECD, Organisation for Economic Cooperation and Development. *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, 1980. Available from: http://www.oecd.org/document/18/0,2340,en_2649342551815186111,00.html, last visit: 12.09. 2004.

Privacy and the National Information Structure. *Principles for Providing and Using Personal Information*, 1995. Available from: http://www.cdt.org/privacy/comments_iitf.html, last visit: September 2004.

Registratiekamer. *Privacy-Enhancing Technologies: The Path to Anonymity*, 1995. Revised edition available from: http://www.cbpweb.nl/documenten/av11Privacy-enhancingtechnologies.htm.

R.S. Rosenberg, *The Social Impact of Computers*. Academic Press, San Diego, 1992.

R.J. Schweizer and H. Burkert. Verwaltungsinformationsrecht. In H. Koller, G. Miller, R. Rhinow and U. Zimmerli, editors, *Schweizerisches Bundesverwaltungsrecht, t. 5, Informations- und Kommunikationsrecht*, pp. 1–45, ed. by R.H. Weber, Helbing & Lichtenhahn, Basel, 1996.

The European Parliament and the Council of the European Union. *EC95 Directive 95/46/EC "On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data"*, Brussels, 1995.

S.D. Warren and L.D. Brandeis. The Right to Privacy. *Harvard Law Review*, 4(5): 193–220, 1890.

A.F. Westin, *Privacy and Freedom*. Atheneum, New York, 1967.